Methodology for the State and County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin (Vintage 2012): April 1, 2010 to July 1, 2012

The U.S. Census Bureau annually produces estimates of the resident population by age, sex, race, and Hispanic origin for each state and county in the United States and the District of Columbia.[1] The following documentation describes the process by which we produce population estimates for these demographic characteristics at the state and county level.

Overview

Resident population includes all residents (both civilian and Armed Forces) living in the United States and is based on the concept of residence used in the 2010 Census that defines a resident of a specified area as a person "usually resident" in that area. The Census Bureau develops estimates by updating the 2010 Census. We begin with modified population counts by age, sex, race, and Hispanic origin from the 2010 Census (see Specification of the Base Population, below) and estimate the change that has occurred since that time. This change is measured annually to produce estimates of the population for July 1, 2010, July 1, 2011, and July 1, 2012. The Vintage 2012 estimates contain the most current data available and supersede all previous estimates. These components are further described in the sections below.

Estimating Population Change

Population can change as a result of births, deaths, or migration, which are known collectively as the components of change. In the United States, births and deaths are recorded with relative accuracy and completeness, and these data are readily available. Migration, on the other hand, can be very difficult to estimate accurately and is the largest source of population change for many areas. For these estimates, migration is divided into two independently estimated sub-components: domestic and international.

We produce separate estimates of the population living in special housing arrangements known as group quarters (for example, college dormitories) because movement into and out of these facilities is unlikely to be captured by our migration estimates, and because we receive data to estimate this population separately. Consequently, our estimation procedure begins by splitting the Census population into two mutually exclusive universes: the group quarters (GQ) population, and the non-GQ or household population. We estimate change in the household population by estimating the components of change mentioned above. Change in the GQ population is estimated using data received annually from members of the Federal-State Cooperative for Population Estimates. The resulting household and GQ estimates are added together to produce the new set of resident population estimates.

Specification of the Base Population

The enumerated resident population from the 2010 Census is the starting point for the post-2010 population estimates. We modify this enumerated population in two ways to produce the April 1, 2010 population estimates base.

First, we reconcile the 2010 Census race categories with the race categories that appear in our administrative data by modifying the "Some other race" responses in the 2010 Census.[2] When a "Some other race" response appears in combination with one or more of the five 1997 Office of Management and Budget (OMB) race categories (White; Black or African American; American Indian and Alaska Native; Asian; and Native Hawaiian and Other Pacific Islander), we drop the "Some other race" response.[3] Some other race alone responses are allocated to one or more of the five OMB categories listed above.

Second, we update the population estimates base to reflect changes to the 2010 Census population due to the Count Question Resolution (CQR) program, legal boundary updates reported by January 1, 2012, and other geographic program revisions.[4]

This editing process produced tabulations from our estimates that show fewer people reporting two or more races than similar tabulations from the 2010 Census, because respondents who selected Some Other Race and one of the OMB mandated races in the 2010 Census appear in the single OMB race category in the estimates base.

In the tables created from these estimates, we group race categories in two different ways. One group includes the five single-race categories and a sixth category that combines all categories with more than one race – referred to in our tables as Two or More Races. The other group includes the five alone-or-in-combination race groups. Each of the alone-or-in-combination groups contains one of the single-race categories plus all the multiple-race categories that include that single race. Alone-or-in-combination groups do not sum to the population total, because each multiple-race person is included in more than one of these groups. For example, people who are White ***and*** Asian would be included in both the White alone-or-in-combination group and the Asian alone-or-in-combination group. We also apply these modifications to the 2010 Census GQ population to produce the GQ estimates base. The GQ estimates base is subtracted from the total estimates base to produce the household estimates base population.

Estimation of the Household Population

The household population is estimated using a technique known as the cohort-component method. In this context, the term *cohort* refers to a group of individuals born in the same time period. The cohort-component method applies the components of population change to groups of individuals based on when they were born. The following equation illustrates how our application of this technique treats annual population change:

$P_1 = P_0 + B - D + NDM + NIM$
where:

P_1 = population at the end of the year
P_0 = population at the beginning of the year
B = births during the year
D = deaths during the year
NDM = net domestic migration during the year
NIM = net international migration during the year

We apply this equation to our beginning population by single year of age, with the result that the population measured by P_1 is always one year older than the population measured by P_0. To produce estimates of the 2012 household population, this technique is repeated for each year following 2010. We begin with an estimate of the July 1, 2010 household population (as described below) and apply the components of change for July 1, 2010 through June 30, 2011 to produce an estimate of the July 1, 2011 household population. We then apply the components of change for July 1, 2011 through June 30, 2012 to the July 1, 2011 population to produce an estimate of the July 1, 2012 population.

1. Estimation of the July 1, 2010 Population

Annual population estimates refer to the midpoint of the year (July 1). The first step in this estimation process is to use the April 1, 2010 household estimates base to develop estimates for July 1, 2010. We do this by controlling the household estimates base to previously produced estimates of the July 1, 2010 household population for higher levels of geography. For the state-level characteristics estimates we control to the July 1 state total estimates and July 1 national estimates by characteristics, and for the county-level estimates we control to the July 1 county total estimates and the July 1 state estimates by characteristics. This is done using the process described below in the section entitled, Ensuring Consistency with Other Estimates.

2. Estimation of Births and Deaths

The birth and death components are estimated using vital records data from two sources. Members of the Federal-State Cooperative for Population Estimates (FSCPE) provide summary data on all registered births and deaths to residents of their respective states by county for calendar years 2010 and 2011. The National Center for Health Statistics (NCHS) provides birth and death data that include sex, race, Hispanic origin, age, and month of occurrence detail for 2010, and total registered births and deaths in 2011. The 2010-2011 county totals from the FSCPE data are controlled to the national total from the NCHS data for the corresponding year, and the 2010 FSCPE data are given the county-level age-sex-race-Hispanic origin distribution from the NCHS data.

For vintage 2012 we receive no data for 2012 and only partial data for 2011, so we create projections to complete our time series. These projections are obtained by applying county-level age-specific fertility and mortality rates to county

population projections for 2011 and 2012. The result is a complete county-level time series of births and deaths for calendar years 2010-2012. We obtain state-level data by summing the county-level data. In order to produce estimates for July 1 of each year, we require components of change for July 1-June 30 intervals, which we refer to as *estimates years*. Calendar year data are converted into estimates years using NCHS month-of-occurrence information, and controlled to the corresponding data from the national estimates time series. We make no adjustments for undercoverage or differential coverage by state, sex, race, Hispanic origin, or age (for deaths).

Data from NCHS on births and deaths differ from inputs that we receive from other agencies in that they are still provided in the four single-race categories specified by OMB's 1977 directive.[5] Since 2003, NCHS has received birth and death data by the revised OMB categories, but not from all states. Consequently, data collected under the old definitions must be converted into the new race categories using race-bridging factors. In the case of births, race-bridging factors are used to first convert the single-race of the mothers and fathers in the birth data to the revised OMB categories.[6] Then, data from the 2010 Census on the race reported for children when the parents are of different races are used to obtain the race of each birth based on the revised race of the mother and father.

3. Estimation of Domestic Migration

We produce estimates of domestic migration at the state and county levels by essentially the same method. This method utilizes data from two sources: annual person-level data from tax returns provided by the Internal Revenue Service (IRS); and the Census Bureau's Demographic Characteristics File (DCF). The DCF is derived from the Social Security Administration's 100 percent file, other administrative records data sources, and the 2010 Census. Keeping in mind that we estimate components of change for *estimates years* that begin with July 1 of one year and continue to June 30 of the next, the first step is to match the person-level IRS data for the two years in question. These matched records contain the addresses from which the returns were filed in both years. The specific dates to which the addresses pertain depend on when the respective tax returns were filed, and may vary from record to record. However, we assume that this information may be used to estimate migration between July 1 of the first year and June 30 of the second.

The matched person-records are then matched to the DCF, which enables us to identify the age, sex, race, and Hispanic origin for each individual.[7] We then tabulate the matched records by these characteristics, place (i.e., state or county) of residence in the first year, and place of residence in the second year. For each county, person-records are classified as *out-migrants* if the first-year address is in that county and the second-year address is in a different county. Similarly, person-records are classified as *in-migrants* if the second-year address is in that place and the first-year address is in a different county.

We use person-records to calculate migration rates and proportions, and we assume they can be applied to the full household population to produce migration estimates even though the tax filers and their dependents do not represent the entire population. An out-migration rate for a given county can be calculated using these data by taking the ratio of the out-migrant records to the total matched records for that county. Multiplying this out-migration rate by an estimate of the household population for that county will produce an estimate of that county's domestic out-migration. We calculate and apply out-migration rates for each county by race, sex, Hispanic origin, and age. These rates are applied to estimates of the household population during the cohort-component process to produce estimates of domestic out-migration for each county by age, sex, race, and Hispanic origin.

In-migration is estimated by allocating out-migration to destination counties using migration in-proportions. Like the migration rates, the migration proportions are computed as the ratio of two sets of person-records. The numerator of this ratio is the sum of the in-migration records for the county in question and the denominator is the sum of the in-migration records for all counties. These in-proportions are computed for all counties by race, sex, Hispanic origin, and age. During the cohort-component process these proportions are applied to the national sum of out-migration by age, race, sex, and Hispanic origin to produce estimates of domestic in-migration for each county.

4. Estimation of Net International Migration

We estimate international migration in several parts: immigration of the foreign born, emigration of the foreign born, net migration between the United States and Puerto Rico, net migration of natives to and from the United States, and net movement of the Armed Forces population to and from the United States. For each component, we first estimate the total migration flow for the nation. To determine the state- and county-level age, sex, race, and Hispanic origin distribution of each component, proxy universes are developed that are assumed to be representative of the different components. The demographic characteristics and geographic distribution of these proxy universes are then applied to the totals for each component. For all components except net movement of the Armed Forces population to and from the United States, state-level characteristics are based on the American Community Survey (ACS) three-year 2008-2010 file. County-level characteristics are based on data from the ACS five-year 2006-2010 file. County-level data are controlled to state-level data to ensure the component data sum as required. For the net movement of the Armed Forces population, demographic characteristics and state distributions are based on data collected by the Defense Manpower Data Center (DMDC) and Census 2000.

Immigration of the foreign born is estimated separately for Mexico and "All other countries" using the ACS question on residence in the prior year. The foreign-born household population who indicated that they lived in Mexico in the prior

year are considered immigrants from Mexico and the foreign-born household population who indicated that they lived abroad but not in Mexico are immigrants from "All other countries." The number of foreign-born migrants who entered the United States between April 2010 and June 2010, for both Mexico and "All other countries," is estimated as one quarter of the foreign-born household population in the 2010 ACS who reported living abroad one year ago. Because this question is asked only of those aged one and older, the estimate of foreign-born immigrants under the age of one is assumed to be equal to half the number of immigrants age one. Information from the 2011 ACS is used to estimate migration for the July 2010 through June 2011 period. This estimate is held constant for the July 2011 through June 2012 period because more recent data are not available. The foreign-born household population whose place of birth was Mexico and whose year of entry was within five years of the survey year is used as the proxy universe to estimate the state- and county-level characteristics of foreign-born immigration from Mexico. The foreign-born household population whose place of birth was other than Mexico and whose year of entry was within five years of the survey year is used as the proxy universe to estimate the state- and county-level characteristics of foreign-born immigration from "All other countries." Age in the ACS is modified for foreign-born immigrants to represent age at arrival to the United States.

Emigration of the foreign born from the United States is estimated separately for Mexico and "All other countries" using a residual method. For foreign-born emigration to Mexico, the foreign-born household population in Census 2000 whose place of birth was Mexico is aged forward (using NCHS life tables) to obtain the expected population in 2007, 2008, 2009, 2010, and 2011. The expected population is then compared to the population estimated in ACS 2007, ACS 2008, ACS 2009, ACS 2010, and ACS 2011. Subtracting the estimated from the expected populations provides us with the residual, which serves as the basis for our emigration rates for the 2000 to 2007, 2000 to 2008, 2000 to 2009, 2000 to 2010, and 2000 to 2011 time periods. This calculation is performed for two period-of-entry groups: the foreign born who entered the United States between 1990 and 1999, and the foreign born who entered before 1990. The method for estimating foreign-born emigration to "All other countries" is the same, except for the foreign-born population whose place of birth was in a country other than Mexico is used in the residual calculations.

We then calculate three-year average rates for each period of entry group and apply the rates to the population at risk of emigrating each year to obtain estimates of emigration of the foreign-born population who entered the United States within the last ten years and of those who entered more than ten years ago. To produce estimates of foreign-born emigration from April 2010 through June 2010, the average of the rates from the 2000 to 2007, 2000 to 2008, and 2000 to 2009 residuals are applied by period of entry to ACS 2009. The estimates are divided by four to obtain estimates for the three-month period. For the estimates from July 2010 through June 2011, the average of the rates from the 2000 to

2008, 2000 to 2009, and 2000 to 2010 residuals are applied to ACS 2010. For the estimates from July 2011 through June 2012, the average of the rates from the 2000 to 2009, 2000 to 2010, and 2000 to 2011 residuals are applied to ACS 2011. The proxy universe for foreign-born emigrants to Mexico who entered the United States within ten years of the estimate year is the foreign-born household population in the ACS whose place of birth was Mexico and who entered the United States within ten years of the survey year.

The proxy universe for foreign-born emigrants to Mexico who entered the United States more than ten years before the estimate year is the foreign-born household population in the ACS whose place of birth was Mexico and who entered the United States more than ten years before the survey year. The proxy universes for the foreign-born emigration to "All other countries" component is the same, except place of birth is restricted to countries other than Mexico.

Data from the ACS and the Puerto Rico Community Survey (PRCS) allow us to estimate the annual migration flows between the United States and Puerto Rico directly using the question on place of prior residence.[8] People who indicated on the ACS that they lived in Puerto Rico one year ago are considered immigrants. People who indicated on the PRCS that they lived in the United States one year ago are considered emigrants. The proxy universe for the net migration between the United States and Puerto Rico is the population born in Puerto Rico whose year of entry was ten or fewer years before the survey year.

The net migration of natives is based on research by Schachter (2008) using data from over 80 countries.[9] This work compared estimates of the U.S. born or U.S. citizen population living overseas measured at two consecutive time periods and used the difference to develop estimates of net native migration. The proxy universe for the net native migration component is the native-born civilian population whose residence one year ago was either in a different state or abroad.

We derive the estimate of the net overseas movement of the Armed Forces population from data collected by DMDC. DMDC provides monthly tabulations of military personnel stationed or deployed outside the United States by age, sex, Hispanic origin, and individual branches of service within the Department of Defense. We assume that change in the overseas military population, excluding deaths, indicates movement of personnel in and out of the United States. To derive the estimates of net movement by race, we apply the race and geographic distribution of the active-duty military population from Census 2000 to DMDC estimates by age, sex, Hispanic origin, and branch of service.

Estimation of the Group Quarters Population

Group Quarters (GQ) population is estimated separately from the household population because of the unique character of this population and our ability to acquire direct data that reflect change in this population. The technique for estimating the GQ population

begins with the GQ base population derived from the 2010 Census. The next step is to estimate GQ change using data supplied by FSCPE members. FSCPE representatives provide independent lists of GQ facilities in their respective states at the county level with the populations typically associated with them at the time of the 2010 Census. They also provide annual updates to this list that we use to calculate the change in the GQ by type of GQ facility. This change is applied to the GQ base to derive annual estimates of the total GQ by type for each county. In states where no GQ data are submitted by the FSCPE, we hold the GQ base data constant. Finally, we distribute these totals by age, sex, race, and Hispanic origin using the distribution of the GQ population by seven major types from the GQ base.

Ensuring Consistency with Other Estimates

The Census Bureau produces a variety of population estimates for different levels of geography and in differing degrees of demographic detail. There can be minor inconsistencies among them because these different estimates utilize different data and processing techniques. For example, when the initial state characteristics estimates are summed to state totals, these totals may differ slightly from the estimates produced by our state totals estimation process. Consequently, the final step in estimates production is to control the estimates to previously produced estimates to ensure consistency. We do this by a technique called *raking*, which involves calculating a *rake factor* as the control total divided by the sum of the numbers we wish to control and then multiplying the numbers we wish to control by the rake factor. In the case of the example just mentioned, we would calculate a rake factor for each state and the District of Columbia and then multiply each area's characteristics estimates by their respective rake factor. This process would produce a set of state characteristics estimates whose totals were consistent with the state totals estimates, but it is likely that many of the new estimates would not be integers. Thus, the final step in this process is to apply a technique we refer to as *controlled rounding*, which enables us to convert the estimates to integers without changing the totals.

The state characteristics estimates must be consistent with both the state totals estimates and the national characteristics estimates. The existence of two independent sets of controls is a complication because raking to one set of controls can upset the consistency with the other set of controls. However, by raking first to one set of controls and then to the other for five iterations, the results are approximately consistent with both sets. A specialized rounding procedure is then applied to maintain consistency with two independent sets of controls.

The situation for county characteristics estimates is similar to that for state characteristics estimates. The county characteristics estimates must be consistent with the county totals estimates and the state characteristics estimates. We accomplish this by iterative raking and our specialized rounding, in the same fashion as we do for the state characteristics estimates. By making the county characteristics estimates consistent with the state characteristics estimates, county characteristics estimates become consistent with the state totals and national characteristics, because the state characteristics are consistent

with these estimates. Thus, by controlling the state characteristics estimates to the state totals and national characteristics, and then controlling the county characteristics estimates to the county totals and state characteristics, we ensure consistency among all these estimates.

---

[1] Throughout this document, the term *county* includes county-equivalents such as parishes and independent cities.

[2] This modification is used for all Census Bureau estimates products and is explained in the document entitled "Modified Race Data Summary File Technical Documentation and ASCII Layout" that can be found on the Census Bureau website at http://www.census.gov/popest/data/historical/files/MRSF-01-US1.pdf. For more information on the procedures used to make these modifications, see http://www.census.gov/popest/research/modified.html.

[3] Office of Management and Budget. *Revisions to the standards for the classification of Federal data on race and ethnicity.* Federal Register 62FR58781-58790, October 30, 1997. Available from: http://www.whitehouse.gov/omb/fedreg/1997standards.html.

[4] For more information on the 2010 Census Count Question Resolution (CQR) program, see http://2010.census.gov/2010census/about/cqr.php.

[5] Office of Management and Budget. *Race and ethnic standards for Federal statistics and administrative reporting*. Statistical Policy Directive 15. May 12, 1977.

[6] For a description of the development of NCHS's race-bridging factors, see: Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, Madans JH. *United States Census 2000 population with bridged race categories.* National Center for Health Statistics. Vital Health Stat 2(135). 2003.

[7] Age is calculated as of the start of the estimation interval using date of birth information from the PCF file.

[8] The Puerto Rico Community Survey was first fielded in 2005. See http://www.census.gov/acs/www/SBasics/FlyerPR.htm for more information.

[9] Schachter, Jason. 2008. "Estimating Native Emigration from the United States," Memorandum.